Iniziato	Thursday, 13 January 2022, 15:14
Stato	Completato
Terminato	Thursday, 13 January 2022, 15:31
Tempo impiegato	17 min. 14 secondi
Punteggio	15,00/15,00
Valutazione	**30,00** su un massimo di 30,00 (**100**%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main reason for the *standardization* of numeric attributes?

Scegli un'alternativa:

a. Map all the numeric attributes to a new range such that the mean is zero and the variance is one.

b. Remove non-standard values

c. Change the distribution of the numeric attributes, in order to obtain gaussian distributions

d. Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following *is not* an objective of feature selection

Scegli un'alternativa:

a. Avoid the *curse of dimensionality*

b. Select the features with higher range, which have more influence on the computations

c. Reduce the effect of noise

d. Reduce time and memory complexity of the mining algorithms

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following statements is *true*?

Scegli una o più alternative:

a. The data which are similar to the majority are never noise

b. The noise always generate outliers

c. The noise can generate outliers

d. Outliers can be due to noise

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In which mining activity the *Information Gain can be useful?*

Scegli un'alternativa:

a. Clustering

b. Classification

c. Discretization

d. Discovery of association rules

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the *Gini Index?*

Scegli un'alternativa:

a. A measure of the *entropy* of a dataset

b. An accuracy measure of a dataset alternative to the *Information Gain* and to the *Misclassification Index*

c. An impurity measure of a dataset alternative to *overfitting* and *underfitting*

d. An impurity measure of a dataset alternative to the *Information Gain* and to the *Misclassification Index*

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a decision tree, an attribute which is used only in nodes near the leaves...

Scegli un'alternativa:

a. ...is irrelevant with respect to the target

b. ...gives little insight with respect to the target

c. ...guarantees high increment of purity

d. ...has a high correlation with respect to the target

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main purpose of *smoothing* in Bayesian classification?

Scegli un'alternativa:

a. Classifying an object containing attribute values which are missing from some classes in the training set

b. Classifying an object containing attribute values which are missing from some classes in the test set

c. Reduce the variability of the data

d. Dealing with missing values

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

With reference to the total *sum of squared errors* and *separation* of a clustering scheme, which of the statements below is true?

a. They are strictly correlated, if, changing the clustering scheme, one increases, then the other does the same
b. They are strictly correlated, if, changing the clustering scheme, one increases, then the other decreases
c. They are two ways to measure the same thing
d. It is possible to optimise them (i.e. minimise SSE and maximise SSB) separately

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What does K-means try to minimise?

Scegli un'alternativa:

a. The *separation*, that is the sum of the squared distances of each cluster centroid with respect tho the global centroid of the dataset
b. The *separation*, that is the sum of the squared distances of each point with respect to its centroid
c. The *distortion*, that is the sum of the squared distances of each point with respect to the points of the other clusters
d. The *distortion*, that is the sum of the squared distances of each point with respect to its centroid

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below is true? (One or more)

Scegli una o più alternative:

a. Sometimes DBSCAN stops to a configuration which does not include any cluster

b. DBSCAN always stops to a configuration which gives the optimal number of clusters

c. Increasing the radius of the neighbourhood can decrease the number of noise points

d. DBSCAN can give good performance when clusters have concavities

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Match the rule evaluation formulas with their names

$$\frac{conf(A \Rightarrow C)}{sup(C)}$$

$$\frac{sup(A \Rightarrow C)}{sup(A)}$$

$$\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$$

$$sup(A \cup C) - sup(A)sup(C)$$

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Vai a...

Consider the transactional dataset below

Machine Learning - Python Lab ►

ID	Items
1	A,B,C
2	A,B,D
3	B,D,E
4	C,D
5	A,C,D,E

Which is the *confidence* of the rule A,C ⇒ B?

Scegli un'alternativa:

a. 50%

b. 40%

c. 100%

d. 20%

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a dataset with D attributes, how many subsets of attributes should be considered for feature selection according to an exhaustive search?

Scegli un'alternativa:

a. O(D)

b. $O(D^2)$

c. $O(2^D)$

d. O(D!)

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

How can we measure the quality of a trained regression model?

a. With a confusion matrix

b. Counting the number of values correctly forecast

c. With precision, recall and accuracy

d. With a formula elaborating the difference between the forecast values and the true ones

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is different from the others?

Scegli un'alternativa:

a. Apriori

b. Decision Tree

c. K-means

d. Expectation Maximisation